NRL Report 8704

Which is the Better Entropy Expression for Speech Processing: $-S \log S$ or $\log S$?

RODNEY JOHNSON AND JOHN E. SHORE

Computer Science and Systems Branch
Information Technology Division

July 20, 1983

NAVAL RESEARCH LABORATORY
Washington, D.C.

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: NRL Report 8704	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE *(and Subtitle)*: WHICH IS THE BETTER ENTROPY EXPRESSION FOR SPEECH PROCESSING: $-S$ LOG S OR LOG S?	5. TYPE OF REPORT & PERIOD COVERED: Interim report on continuing NRL problem
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Rodney Johnson and John E. Shore	8. CONTRACT OR GRANT NUMBER(s): 61153N, RR-21-05-42, 75-0107-0-3
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Naval Research Laboratory, Washington, DC 20375	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Reserach, Arlington, VA 22217	12. REPORT DATE: July 20, 1983
	13. NUMBER OF PAGES: 21
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*	15. SECURITY CLASS. *(of this report)*: UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Information theory
Maximum entropy
Minimum cross-entropy
Spectrum estimation
Speech processing

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

In maximum-entropy spectral analysis (MESA), one maximizes the integral of $\log S(f)$, where $S(f)$ is a power spectrum. The resulting spectral estimate, which is equivalent to that obtained by linear prediction and other methods, is popular in speech-processing applications. An alternative expression, $-S(f) \log S(f)$, is used in optical processing and elsewhere. This report considers whether the alternative expression leads to spectral estimates useful in speech

(Continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

20. ABSTRACT (Continued)

processing. We investigate the question both theoretically and empirically. The theoretical investigation is based on generalizations of the two estimates—the generalizations take into account prior estimates of the unknown power spectrum. It is shown that both estimates result from applying a generalized version of the principle of maximum entropy, but they differ concerning the quantities that are treated as random variables. The empirical investigation is based on speech synthesized using the different spectral estimates. Although both estimates lead to intelligible speech, speech based on the MESA estimate is qualitatively superior.

CONTENTS

WHICH IS THE BETTER ENTROPY EXPRESSION FOR SPEECH PROCESSING: $-S \log S$ OR $\log S$?

INTRODUCTION

Because the power spectrum $S(f)$ of a band-limited stationary process is related to its autocorrelation function $R(t)$ by a Fourier transform, and because it is relatively easy to measure $R(t)$, many spectral analysis techniques start with values of $R(t)$. If $R(t)$ is known at a set of points or regions, one class of spectral analysis techniques proceeds by extrapolating $R(t)$ so as to take on reasonable values in the unknown regions. The extrapolated autocorrelation function is equivalent to a power-spectrum estimate by a Fourier transform.

Perhaps the best known extrapolation technique is Burg's maximum-entropy spectral analysis (MESA) [1,2], in which the power spectrum $S(f)$ is estimated by maximizing

$$\int_0^W \log S(f)\, df \tag{1}$$

subject to the constraints

$$R_r = R(t_r) = \int_{-W}^{+W} S(f) \exp(2\pi i t_r f)\, df, \tag{2}$$

where W is the bandwidth and where $R(t_r)$, $r = 1, 2, \ldots, M$, are known values of the autocorrelation function. The MESA estimate of $S(f)$ has the well-known all-pole, autoregressive, or linear prediction form, which can also be derived by various equivalent formulations [3-6]. It has become one of the most widely used spectral analysis techniques in geophysical data processing [7-9] and speech processing [4,10].

"Maximum-entropy spectral analysis" is also used in image processing. In that field, however, the phrase refers not only to successful estimates produced by maximizing (1) [11-13], but also to estimates produced by maximizing [14-16]

$$-\int_0^W S(f) \log S(f)\, df. \tag{3}$$

Spectral estimates based on (3) have also been studied for ARMA and meteorological time series [17,18]. Although there is controversy in the image-processing literature about whether (1) or (3) yields the best estimates [16,19], the success of (3) in image processing raises the question of whether (3) might also be useful in speech processing. We consider the question in this report and attempt to answer it. As part of our investigation, we also derive a generalization of the estimate produced by maximizing (3), one that takes into account a prior estimate of the unknown power spectrum.

Our report is organized as follows: In the next section we review derivations of the forms (1) and (3), and we discuss theoretical arguments for each of them. We then turn to an empirical comparison. Our approach is discussed in the third section, and the results are summarized in the fourth section. A brief general discussion then follows in the concluding section.

Manuscript approved February 14, 1983.

BACKGROUND

In this section we derive the different spectral estimators that result from maximizing (1) and (3). We find that they both result from applying a generalized form of the principle of maximum entropy [20-22], but they differ concerning the quantities that are treated as random variables. In the case of (1), the underlying random variables are the coefficients of a Fourier-series model, and the spectral powers $S(f)$ are expected values. In the case of (3), the spectral power $S(f)$—suitably normalized—is treated as a probability density, and the underlying random variable is the frequency.

The $-\log S$ Form

In deriving MESA, Burg's approach was to extrapolate $R(t)$ in a manner that maximizes the entropy of the underlying stochastic process [1,2]. This is an application of the principle of maximum entropy [20-22]. An intuitive justification for such an extrapolation of $R(t)$ is that it agrees with what is known—as expressed by the constraints (2)—while being "maximally noncommittal" about what is not known [20]. In particular, (1) is the entropy gain in a stochastic process that is passed through a linear filter with characteristic function $Y(f)$, where $S(f) = |Y(f)|^2$, as described in Refs. 9 (pp. 412-414), 23 (pp. 93-95), and 24 (p. 243). If the input process is white noise, then the output process has spectral power density $S(f)$. This suggests that the process entropy can be maximized by maximizing the entropy gain of the filter that produces the process. Thus (1) is maximized subject to the constraints (2). The result is

$$S(f) = \frac{\sigma^2}{\left|\sum_{r=0}^{M} a_r z^{-r}\right|^2}, \tag{4}$$

where $z = \exp(-2\pi i f \Delta t)$. This is the familiar MESA [2] or linear-prediction-coding (LPC) [4] estimate. The a_r are the inverse-filter sample coefficients, and σ^2 is the gain. Such derivations of (4) have several mathematical and logical drawbacks [25]. For example, entropy is mathematically ill-behaved for continuous densities [26, pp.31-32]. A derivation of MESA without these drawbacks arises as a special case of minimum cross-entropy spectral analysis (MCESA) [25] and also helps to expose the difference underlying the choice of maximizing (1) or (3).

Like MESA, MCESA is an information-theoretic extrapolation of $R(t)$, but it differs from MESA in that it accounts for a prior estimate of $S(f)$ (or $R(t)$). MCESA is based on the principle of minimum cross-entropy (discrimination information, directed divergence, Kullback-Leibler number, relative entropy) [27-30]. Cross-entropy minimization estimates an unknown probability density $q^{\dagger}(\mathbf{x})$ from a prior estimate $p(\mathbf{x})$ and known expected values

$$\int q^{\dagger}(\mathbf{x}) g_r(\mathbf{x})\, d\mathbf{x} = \overline{g}_r\,, \tag{5}$$

where $r = 0, \ldots, M$. The estimate is obtained by minimizing the cross-entropy

$$H(q,p) = \int q(\mathbf{x}) \log\left[\frac{q(\mathbf{x})}{p(\mathbf{x})}\right] d\mathbf{x} \tag{6}$$

subject to the constraints (5) and

$$\int q(\mathbf{x})\, d\mathbf{x} = 1. \tag{7}$$

When $p(\mathbf{x})$ is interpreted as a prior estimate, cross-entropy minimization can be viewed as a generalization of entropy maximization [28]—cross-entropy minimization reduces to entropy maximization when $p(\mathbf{x})$ is uniform. When $p(\mathbf{x})$ is interpreted as an "invariant measure" as in [30], the two principles can be viewed as equivalent. In either case, the resulting estimate of $q^{\dagger}(\mathbf{x})$ has the form [27,29,31]

$$q(\mathbf{x}) = p(\mathbf{x}) \exp\left[-\lambda - \sum_{r=0}^{M} \beta_r f_r(\mathbf{x})\right], \tag{8}$$

where the β_r and λ are Lagrangian multipliers determined by (5) and (7). We refer to $p(\mathbf{x})$ as a *prior.*

In deriving MCESA we consider time-domain signals of the form

$$s(t) = \sum_{k=1}^{N} a_k \cos(2\pi f_k t) + b_k \sin(2\pi f_k t), \tag{9}$$

where the a_k and b_k are random variables and where the f_k are nonzero frequencies. Since any stationary random process $g(t)$ can be obtained as the limit of a sequence of processes with discrete spectra [32, p. 36], (9) is quite general. With suitable choices for the frequencies and amplitudes, the mean square error $E(|g(t)-s(t)|^2)$ can be made arbitrarily small. Since the power at frequency f_k is $x_k \equiv \frac{1}{2}(a_k^2+b_k^2)$, we describe the random process in terms of a joint probability density $q(\mathbf{x})$, where $\mathbf{x} = x_1, x_2, \ldots, x_N$.

Let $S_k^\dagger$ be the spectral power at frequency f_k of some unknown process $q^\dagger(\mathbf{x})$:

$$S_k^\dagger = \int x_k q^\dagger(\mathbf{x})\, d\mathbf{x}. \tag{10}$$

Furthermore let P_k be a prior estimate of $S_k^\dagger$. As a form for the prior estimate of the probability density $q^\dagger$, we assume

$$p(\mathbf{x}) = \prod_{k=1}^{N} \frac{1}{P_k} \exp\left[\frac{1}{P_k}\right]. \tag{11}$$

This assumption is consistent with the prior spectral-power estimates, since $\int x_k p(\mathbf{x})\, d\mathbf{x} = P_k$, and it is equivalent to a Gaussian prior assumption for the amplitudes a_k and b_k in (9) [25]. Suppose that one obtains new information about $q^\dagger$ in the form of $M+1$ values of the autocorrelation function $R(t_r)$:

$$\begin{aligned} R_r = R(t_r) &= \sum_{k=-N}^{+N} S_k^\dagger \exp(2\pi i t_r f_k) \\ &= \sum_{k=1}^{N} 2S_k^\dagger \cos(2\pi t_r f_k), \end{aligned} \tag{12}$$

where $r = 0, \ldots, M$, and $t_0 = 0$. Using (10), we write this as

$$R_r = \int \left[\sum_{k=1}^{N} 2x_k \cos(2\pi t_r f_k)\right] q^\dagger(\mathbf{x})\, d\mathbf{x}, \tag{13}$$

which has the form of expected-value constraints (5). Given the prior (11) and the constraints (13), one can compute a minimum cross-entropy posterior estimate $q(\mathbf{x})$ of the form (8). The result can be written [25] as

$$q(\mathbf{x}) = \prod_{k=1}^{N} \left[\frac{1}{P_k} + u_k\right] \exp\left[-\left(\frac{1}{P_k} + u_k\right) x_k\right], \tag{14}$$

where

$$u_k = \sum_{r=0}^{M} 2\beta_r \cos(2\pi t_r f_k).$$

The β_r are Lagrangian multipliers determined by the constraints (13). The posterior estimate of the power spectrum is just $S_k = \int x_k q(\mathbf{x})\, d\mathbf{x}$, which becomes

$$S_k = \frac{1}{\dfrac{1}{P_k} + \displaystyle\sum_{r=0}^{M} 2\beta_r \cos(2\pi t_r f_k)}, \tag{15}$$

where the β_k are chosen so that the S_k satisfy the autocorrelation constraints

$$R_r = \sum_{k=1}^{N} 2S_k \cos(2\pi t_r f_k). \tag{16}$$

If one assumes a flat prior estimate of the prior spectrum, $P_k = P$, and equal spacing of the autocorrelation lags, $t_r = r\Delta t$, (15) can be written in the form (4) [25].

From the foregoing we see that the form (4) results from treating the Fourier power variables $x_k = \frac{1}{2}(a_k^2 + b_k^2)$ as random variables and applying cross-entropy minimization to the probability density $q(\mathbf{x})$. The same results are obtained if one treats the Fourier amplitudes a_k and b_k themselves as the random variables [25,33]. To see the relationship between the maximization form (1), the spectral estimate (4), and the underlying density $q(\mathbf{x})$ more directly, note that the posterior probability density (14) can be expressed in terms of the posterior spectral-power estimates (15):

$$q(\mathbf{x}) = \prod_{k=1}^{N} \frac{1}{S_k} \exp\left[-\frac{x_k}{S_k}\right]. \tag{17}$$

Computing the normalized differential entropy of the posterior power estimates (15) yields

$$-\frac{1}{N}\int q(\mathbf{x}) \log q(\mathbf{x})\, d\mathbf{x} = 1 + \frac{1}{N}\sum_{k=1}^{N} \log S_k. \tag{18}$$

Except for the constant, which has no effect on maximization, the right-hand side of (18) is the discrete form of (1). Maximizing (18) subject to the constraints (16) leads again to (4).

The −S log S Form

Results from the preceding subsection show that the MESA or LPC spectral estimate (4) is the result of applying maximum entropy or minimum cross-entropy with the coefficients of the underlying Fourier series (9) treated as random variables. In arguing for the maximization of $-\sum_k S_k \log S_k$ rather than $\sum_k \log S_k$. Skilling [16] points out that the goal is to estimate the power spectrum itself, not the Fourier amplitudes in an underlying model like (9), so that a more direct and better estimate should result from treating the unknown power-spectrum variables $S_k^\dagger$ as probabilities. Mathematically this is reasonable, provided that the power spectrum is normalized so that $\sum_k S_k^\dagger = 1$. The difference in the two approaches is illustrated well by (10). In the first approach, one assumes that the $S_k^\dagger$ are expectations of an underlying probability density $q^\dagger(\mathbf{x})$, and one expresses the known autocorrelations as expectations of $q^\dagger(\mathbf{x})$ as in (13); it follows from the preceding subsection that one should maximize $\sum_k \log S_k$. In the second approach, one assumes that the $S_k^\dagger$ are probabilities, and one expresses the known autocorrelations as expectations of the probability distribution $S_k^\dagger$, $k = 1, \ldots, N$, as in (12) (we defer for the moment details concerning correct normalization); it follows that maximum entropy implies the maximization of $-\sum_k S_k \log S_k$.

In deriving the power spectrum estimate that results from maximizing $-\sum_k S_k \log S_k$, we proceed with the general case involving a prior estimate and cross-entropy minimization as in the previous subsection. Since we assumed a known autocorrelation for lag $t_0 = 0$, $\sum_k S_k^\dagger = \frac{1}{2}R_0$ is known. Let $\mathbf{q}^\dagger = \{q_1^\dagger, q_2^\dagger, \ldots, q_N^\dagger\}$ and $\mathbf{p} = \{p_1, p_2, \ldots, p_N\}$ be probability distributions defined by normalizing the power spectra $S_k^\dagger$ and P_k:

$$q_k^\dagger = \frac{2S_k^\dagger}{R_0},$$

$$p_k = \frac{P_k}{T},$$

where P_k is a prior estimate of $S_k^\dagger$, and where

$$T = \sum_{k=1}^{N} P_k.$$

We rewrite the autocorrelation constraints (12) as expectations of $\mathbf{q}^\dagger$:

$$\frac{2R_r}{R_0} = \sum_{k=1}^{N} 2 \cos(2\pi t_r f_k)\, q_k^\dagger. \tag{19}$$

Then we obtain a posterior estimate of $\mathbf{q}^\dagger$ by minimizing the cross-entropy

$$H(\mathbf{q}, \mathbf{p}) = \sum_{k=1}^{N} q_k \log \frac{q_k}{p_k}$$

subject to the constraints (19). Note that the constraint for $r = 0$ reduces to the normalization constraint $\sum_k q_k = 1$. The result is

$$q_k = p_k \exp\left[-\sum_{r=0}^{M} 2\mu_r \cos(2\pi t_r f_k)\right], \tag{20}$$

where the μ_r are chosen to satisfy the constraints. We define the posterior power-spectrum estimate as $S_k = \frac{1}{2} R_0 q_k$, which satisfies (12).

We have

$$\sum_{k=1}^{N} S_k \log \frac{S_k}{P_k} = \tfrac{1}{2} R_0 H(\mathbf{q}, \mathbf{p}) + \tfrac{1}{2} R_0 \log \frac{R_0}{2T},$$

where $\frac{1}{2}R_0$ and $\frac{1}{2}R_0 \log(R_0/2T)$ are constant and $\frac{1}{2}R_0 > 0$. It follows that minimizing $H(\mathbf{q}, \mathbf{p})$ is equivalent to minimizing

$$\sum_{k=1}^{N} S_k \log \frac{S_k}{P_k}. \tag{21}$$

Minimizing (21) subject to the constraints (12) yields

$$S_k = P_k \exp\left[-\sum_{r=0}^{M} 2\phi_r \cos(2\pi t_r f_k)\right], \tag{22}$$

where the ϕ_r are chosen to satisfy the constraints. The ϕ_r in (22) are equal to the μ_r in (20) except for ϕ_0, which satisfies $\phi_0 = \mu_0 + \frac{1}{2} \log(R_0/2T)$.

For a flat prior estimate $P_k = P$, minimizing (21) is equivalent to maximizing

$$-\sum_{k=0}^{N} S_k \log S_k,$$

which is just the discrete form of (3). Spectral estimates based on the minimization of (21) have been reported recently in Ref. 34. Also, a first-order approximation of the estimate (22) appears to be equivalent to the PDFT estimator introduced in Refs. 35 and 36.

Summary

Here we summarize the final results for the two spectral estimates. Both estimates proceed from a prior estimate P_k and known autocorrelations R_r. When the coefficients in an underlying Fourier-series model are treated as random variables and the S_k are treated as expectations, cross-entropy minimization leads to the estimate

$$S_k = \frac{1}{\dfrac{1}{P_k} + \displaystyle\sum_{r=0}^{M} 2\beta_k \cos(2\pi t_r f_k)}. \tag{24}$$

For the case of a flat prior estimate $P_k = P$, (24) follows from maximizing $\sum_k \log S_k$. When the S_k are treated as probabilities rather than expectations, cross-entropy minimization leads to the estimate

$$S_k = P_k \exp\left[-\sum_{r=0}^{M} 2\phi_r \cos(2\pi t_r f_k)\right]. \tag{25}$$

For the case of a flat prior estimate, (25) follows from maximizing $-\sum_k S_k \log S_k$. Because the result in this case arises from performing maximum entropy on a probability distribution defined by normalizing a power spectrum, we refer to it as maximum-entropy normalized spectral analysis (MENSA).**

The Lagrangian multipliers β_r in (24) and ϕ_r in (25) are chosen in both cases so that the estimates agree with the known autocorrelations

$$R_r = \sum_{k=1}^{N} 2S_k \cos(2\pi t_r f_k), \tag{26}$$

where $r = 0, 1, \ldots, M$. Note that, given one of the spectral estimates S_1 through S_N, substitution of an arbitrary lag t for t_r in (26) defines the corresponding extrapolation of the known autocorrelations.

Which of the two estimates (24) and (25) is better? In our opinion, if one has a good physical model for some variable of interest, and if the model can be incorporated into the derivation of an estimate for that variable, it makes sense to do so. Because such estimates can exploit more information than estimates derived without an underlying model, estimates based on underlying models should be better. Furthermore, although we recognize that normalizing the S_k and treating them as probabilities is mathematically sound, we do not see any reasonable physical interpretation. What events have probabilities proportional to S_k? This suggests that (24) is better. Also, since (24) yields all-pole models in the important case of flat priors, since all-pole spectra result from passing a broadband signal through a multilayered transmission medium, and since the human vocal tract is a multilayered transmission medium, it follows that (24) should be appropriate for speech processing. On the other hand, arguments for (25) also have merit, and it is clear that the best method of answering the question is empirical. This we attempt to do in the remainder of this report.

EXPERIMENTAL APPROACH

In this section we present basic definitions, discuss our experimental approach, and discuss various computational issues. Our general approach is to process various speech signals in order to compare measured power spectra and autocorrelations with MESA and MENSA estimates. We also synthesize speech using both MESA and MENSA power-spectrum estimates and qualitatively compare the results.

Definitions and Notation

Let $\mathbf{y} \equiv \{y_1, y_2, \ldots, y_L\}$ comprise L time-domain samples, equispaced at intervals of Δt, from one "frame" of speech data. From $\mathbf{y}$ we compute estimated autocorrelations $\mathbf{R} \equiv \{R_0, R_1, \ldots, R_{L-1}\}$ by means of

$$R_r = \frac{1}{L} \sum_{i=1}^{L-r} y_i y_{i+r}. \tag{27}$$

This is a biased estimate, but it guarantees positive-definiteness. Let $\mathbf{Q} \equiv \{Q_1, Q_2, \ldots, Q_N\}$ be the power spectrum defined by the discrete Fourier transform of the measured autocorrelations. Defining $R_{-r} = R_r$, we have

**This somewhat contrived acronym has the additional virtue of being the Latin word for table, which is the source of the Spanish word for table (mesa).

$$Q_k = \sum_{r=-L+1}^{L-1} R_r \exp(-2\pi i t_r f_k)$$

$$= R_0 + \sum_{r=1}^{L-1} 2R_r \cos(2\pi t_r f_k). \tag{28}$$

As the N discrete frequencies we take

$$f_k = (k - \tfrac{1}{2}) \frac{1}{2N\Delta t}.$$

That is, we divide the interval from zero to the Nyquist frequency into N subintervals and define f_k as the midpoint of each subinterval.

Let $\mathbf{S} \equiv \{S_1, S_2, \ldots, S_N\}$ be the power-spectrum estimate obtained from (24) using a flat prior estimate and the first $M + 1$ autocorrelations R_r from (27). $\mathbf{S}$ is the standard MESA or LPC estimate of the power spectrum—its usual, continuous-frequency form is given by (4). Let $\mathbf{S}^* \equiv \{S_1^*, S_2^*, \ldots, S_N^*\}$ be the MENSA power-spectrum estimate obtained from (25) using the same flat prior estimate and the same $M+1$ autocorrelations from (27). Finally, let $\mathbf{A}$ and $\mathbf{A}^*$ be the extrapolated autocorrelations for all L lags $t_r = r\Delta t$, $r = 0, \ldots, L - 1$, obtained from (26) using $\mathbf{S}$ and $\mathbf{S}^*$ respectively. Note that A_r and A_r^* match the actual autocorrelations (27) for $r = 0, \ldots, M$. For $r > M$, however, A_r and A_r^* are in general different from each other and from R_r. For convenience, we summarize the notation as follows:

$\mathbf{y}$ = a vector of L time-domain samples from one speech frame,

$\mathbf{R}$ = the measured autocorrelations for L lags computed from $\mathbf{y}$,

$\mathbf{Q}$ = the "actual" power spectrum defined by a Fourier transform of $\mathbf{R}$,

$\mathbf{S}$ = a MESA or LPC estimate of the power spectrum from first $M + 1$ lags of $\mathbf{R}$,

$\mathbf{S}^*$ = a MENSA estimate of the power spectrum from first $M + 1$ lags of $\mathbf{R}$,

$\mathbf{A}$ = a MESA or LPC autocorrelation extrapolation based on $\mathbf{S}$,

$\mathbf{A}^*$ = a MENSA autocorrelation extrapolation based on $\mathbf{S}^*$.

For the work reported here, we used $L = 180$ and $M = 8, 10, 25$. When we refer to more than one speech frame, we add a subscript to the foregoing definitions.

What and How to Compare

Much work in speech analysis and synthesis uses $\mathbf{S}$ to model the power spectrum. We are interested in testing the hypothesis that using $\mathbf{S}^*$ would lead to better results. To obtain information that could help to confirm or refute the hypothesis, we did three things: (a) For a variety of representative speech frames we plotted $\mathbf{A}$, $\mathbf{A}^*$, and $\mathbf{R}$ and we performed qualitative and quantitative comparisons. (b) For the same frames we plotted $\mathbf{S}$, $\mathbf{S}^*$, and $\mathbf{Q}$ and performed qualitative comparisons. (c) We performed qualitative comparisons of speech synthesized two different ways: we used identical pitch and voicing decisions and used either $\mathbf{S}$ or $\mathbf{S}^*$ for spectral shape.

What about quantitative comparisons? For some distortion measure d, one could compare $d(\mathbf{Q},\mathbf{S})$ with $d(\mathbf{Q},\mathbf{S}^*)$, but what is the right choice for d? Clearly, one distortion measure could yield $d(\mathbf{Q},\mathbf{S}) < d(\mathbf{Q},\mathbf{S}^*)$, while another could yield the reverse inequality. One reasonable choice is the Itakura-Saito distortion d_{IS} [37], which is known to be useful in speech processing:

$$d_{IS}(\mathbf{Q},\mathbf{S}) = \frac{1}{N}\sum_{k=1}^{N}\left[\frac{Q_k}{S_k} - 1 - \log\frac{Q_k}{S_k}\right].$$

But in the notation of our background section the Itakura-Saito distortion $d_{IS}(\mathbf{S},\mathbf{P})$ is just the asymptotic cross-entropy between $q(\mathbf{x})$ and $p(\mathbf{x})$; hence, derivations of MESA or LPC spectra by cross-entropy minimization are equivalent to derivations by minimization of Itakura-Saito distortions [38,25,10]. Not only does $\mathbf{S}$ minimize $d_{IS}(\mathbf{S},\mathbf{P})$ subject to the constraints, but $\mathbf{S}$ is the spectrum of the form (15) that minimizes $d_{IS}(\mathbf{Q},\mathbf{S})$ [37,39]. Use of the Itakura-Saito distortion might therefore involve an intrinsic bias in favor of MESA. That is not to imply that $d_{IS}(\mathbf{Q},\mathbf{S})$ is necessarily less than $d_{IS}(\mathbf{Q},\mathbf{S}^*)$. However, we wish to avoid relying entirely on a distortion measure that relates specially to the mathematics of one or the other of the two estimates.

We therefore consider a distortion measure that bears a relation to MENSA analogous to that of d_{IS} to MESA. We define the "cross-entropy distortion" $d_{CE}(\mathbf{Q},\mathbf{S})$ to be the cross-entropy of the probability distributions obtained by normalizing $\mathbf{Q}$ and $\mathbf{S}$:

$$d_{CE}(\mathbf{Q},\mathbf{S}) = \sum_{k=1}^{N}\frac{Q_k}{\sum_{j=1}^{N}Q_j}\log\frac{Q_k}{S_k} - \log\frac{\sum_{j=1}^{N}Q_j}{\sum_{j=1}^{N}S_j}.$$

Then $\mathbf{S}^*$ minimizes $d_{CE}(\mathbf{S}^*,\mathbf{P})$ subject to constraints just as $\mathbf{S}$ minimizes $d_{IS}(\mathbf{S},\mathbf{P})$ subject to constraints. Moreover $\mathbf{S}^*$ is one of the spectra of the form (22) that minimizes $d_{CE}(\mathbf{Q},\mathbf{S}^*)$ [29]. We use d_{CE} as well as d_{IS} for quantitative comparisons; that is, we compare $d_{CE}(\mathbf{Q},\mathbf{S})$ with $d_{CE}(\mathbf{Q},\mathbf{S}^*)$.

We also use a third distortion measure, the gain-optimized Itakura-Saito distortion [39] defined by $d_{GO}(\mathbf{Q},\mathbf{S}) = \min_g d_{IS}(g\,\mathbf{Q},\mathbf{S})$, where g ranges over positive constant scale factors. This is closely related to d_{IS} but, like d_{CE}, is insensitive to changes in the gains of the two spectra. It can be computed from

$$d_{GO}(\mathbf{Q},\mathbf{S}) = \log\frac{1}{N}\sum_{k=1}^{N}\frac{Q_k}{S_k} - \frac{1}{N}\sum_{k=1}^{N}\log\frac{Q_k}{S_k}.$$

Numerical Issues and Procedures

The MENSA estimate $\mathbf{S}^*$ can be produced by an algorithm that determines minimum cross-entropy probability distributions given arbitrary priors and arbitrary constraints. The mathematics underlying a Newton-Raphson-based algorithm is discussed in Appendix A of Ref. 29, and an APL program that implements this algorithm is described in detail in Ref. 40. For the work reported here, we used a FORTRAN version of the APL program. The resulting $\mathbf{S}^*$ may be thought of as a discrete-frequency approximation to a continuous power spectrum, one that is defined by the maximization of $-\int S(f)\log S(f)\,df$ rather than $-\sum_k S_k \log S_k$. The accuracy of the discrete-frequency approximation will depend on the number of frequency points N. Although it would better to use an algorithm that produced a continuous representation of the MENSA estimate, we do not have such an algorithm.

As for $\mathbf{S}$, a variety of methods are available. Standard MESA or LPC methods can produce the inverse-filter coefficients used in (4) or any of the equivalent sets of parameters such as reflection coefficients. The result is a continuous representation of the spectral estimate that can then be evaluated at the frequencies f_k in order to yield $\mathbf{S}$. No doubt this is more accurate than a method that computes a discrete-frequency approximation, but to use it might introduce a misleading source of differences between $\mathbf{S}$ and $\mathbf{S}^*$. To avoid such a problem, we chose to compute the $\mathbf{S}$ in a manner analogous to the computation of $\mathbf{S}^*$. In particular we used a FORTRAN implementation of the MCESA [25] algorithm described in Ref. 41. This algorithm uses the Newton-Raphson method to compute (24) for arbitrary priors and autocorrelation constraints. For a flat prior, the result is just a

discrete-frequency approximation to a continuous MESA or LPC spectrum. As checks on the discrete-frequency computations of $\mathbf{S}^*$ and $\mathbf{S}$, we obtained results for various values of the number of frequency points N, and we compared the results for $\mathbf{S}$ with continuous frequency results obtained using Levinson recursion.

In considering how to obtain synthetic speech using the two different spectral shapes, we decided to take advantage of commonly available, LPC-based programs. This approach, which is ideal for MESA spectra, involves exciting an all-pole filter with either white noise, for unvoiced sounds, or a periodic pulse train, for voiced sounds. For MENSA spectra, which do not have the all-pole form, we had to proceed indirectly. Our procedure was as follows: First we analyzed the test sentence for pitch and voicing using a modified cepstral technique described in Ref. 42 and implemented in Version 4.0 of the Interactive Laboratory System (ILS) from Signal Technology, Inc. The results were used for both syntheses. For the synthesis based on $\mathbf{S}^*$ we used a 29th-order all-pole approximation to the power spectrum $\mathbf{S}^*$ in each frame. This approximation was computed by taking the first 29 lags of the autocorrelation extrapolation $\mathbf{A}^*$ and using Levinson recursion to yield a set of reflection coefficients. As checks we plotted the resulting approximate power spectrum and compared it with $\mathbf{A}^*$. For the synthesis based on $\mathbf{S}$ we followed the same procedure—we ran Levinson recursion on the first 29 lags of $\mathbf{A}$. Had we been dealing with exact, continuous spectra, the resulting "approximate" spectrum would be exactly equal to $\mathbf{S}$, so it would have been reasonable to bypass this step. We included it, however, to keep the comparison as fair as possible. As a check we also synthesized speech using spectral shapes obtained directly from Levinson recursion on the first $M + 1$ lags of the measured autocorrelations $\mathbf{R}$. Note that the 29th-order all-pole synthesis spectra are 29th-order approximations to $\mathbf{S}$ and $\mathbf{S}^*$ and not 29th-order approximations to $\mathbf{Q}$.

EXPERIMENTAL RESULTS

We obtained results for the sentence "*The meeting begins at four p.m.*" The sentence was spoken by a male, passed through an antialiasing filter, digitized at 8000 samples per second, and divided into 100 frames of 180 samples each. Using 256 discrete frequencies ($N = 256$), we computed $\mathbf{R}_j$, $\mathbf{Q}_j$, $\mathbf{S}_j^*$, $\mathbf{A}_j^*$, $\mathbf{S}_j$, and $\mathbf{A}_j$, $j = 1, \ldots, 100$, as discussed in the preceding section. We also did computations for some cases with $N = 64$ and $N = 128$. In general there were no essential differences between results for $N = 64$ or 128 and $N = 256$. It is a frequent practice to preprocess speech samples before the autocorrelations are estimated—the y_i in (27) are the result of preemphasizing or windowing the speech samples. We therefore repeated the computations using Hamming windowing alone, 90% preemphasis alone, and both together. In the following we focus attention on two frames: frame 56, which contains a portion of the phoneme /*f*/, and frame 39, which contains a portion of the phoneme /*I*/. For convenience we refer to these frames by means of the subscripts f and I respectively. Unless windowing or preemphasis is explicitly mentioned, the reference is to the spectra computed without preprocessing.

Comparison of Autocorrelation Extrapolations

In Fig. 1 we plot $\mathbf{R}_f$, $\mathbf{A}_f^*$, and $\mathbf{A}_f$ for $N = 256$. When we plotted the continuous autocorrelation function obtained by Levinson recursion, it was indistinguishable from $\mathbf{A}_f$, which implies that the discrete frequency approximations are accurate. Beyond the constraint limit of lag 10, the extrapolations $\mathbf{A}_f^*$ and $\mathbf{A}_f$ differ from each other as well as from $\mathbf{R}$. One would be hard pressed to argue that either one is a "better" extrapolation. The same conclusion follows from Fig. 2, in which we plot analogous results from the frame containing a portion of the phoneme /*I*/.

Comparison of Power Spectra

Turning to the power spectra, we plot $\mathbf{S}_f^*$, $\mathbf{S}_f$, $\mathbf{S}_I^*$, and $\mathbf{S}_I$ in Figs. 3 through 6 for $M = 10$. The spectra $\mathbf{S}_f^*$ and $\mathbf{S}_f$ are quite similar; $\mathbf{S}_I^*$ and $\mathbf{S}_I$ are quite different. In particular, $\mathbf{S}_I^*$ has deep nulls that

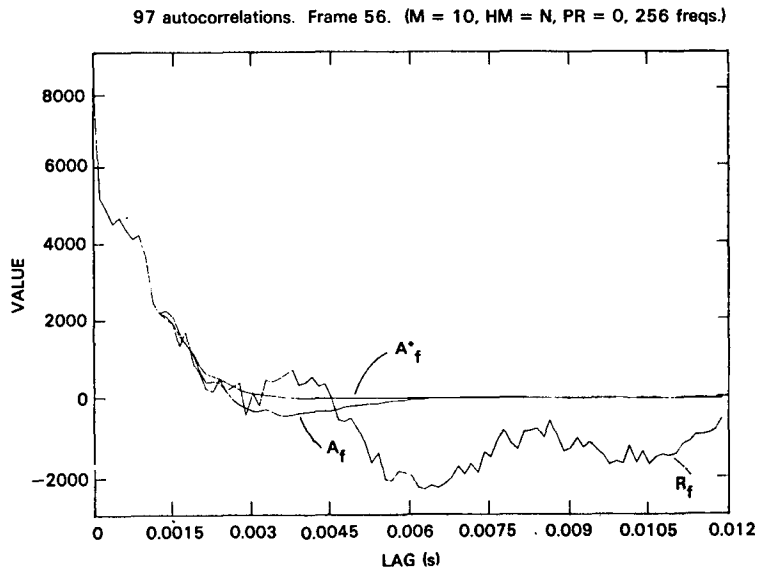

Fig. 1 — Autocorrelations from speech samples and from MESA and MENSA spectral estimates (/*f*/)

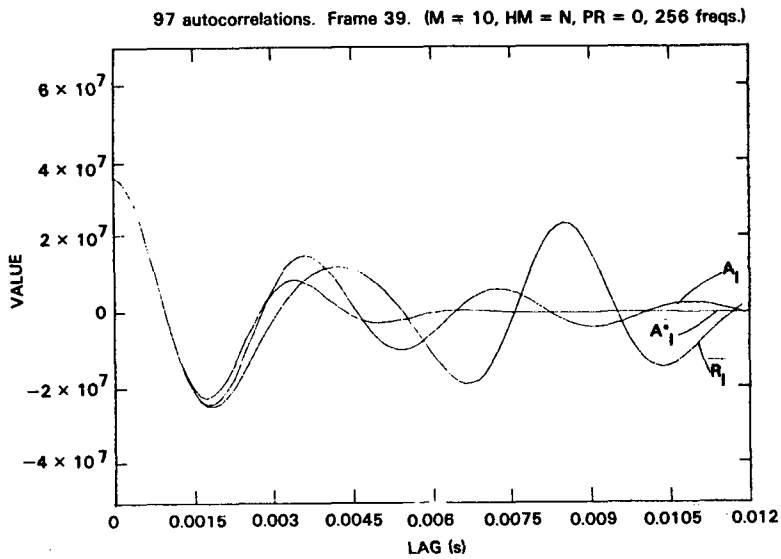

Fig. 2 — Autocorrelations from speech samples and from MESA and MENSA spectral estimates (/*I*/)

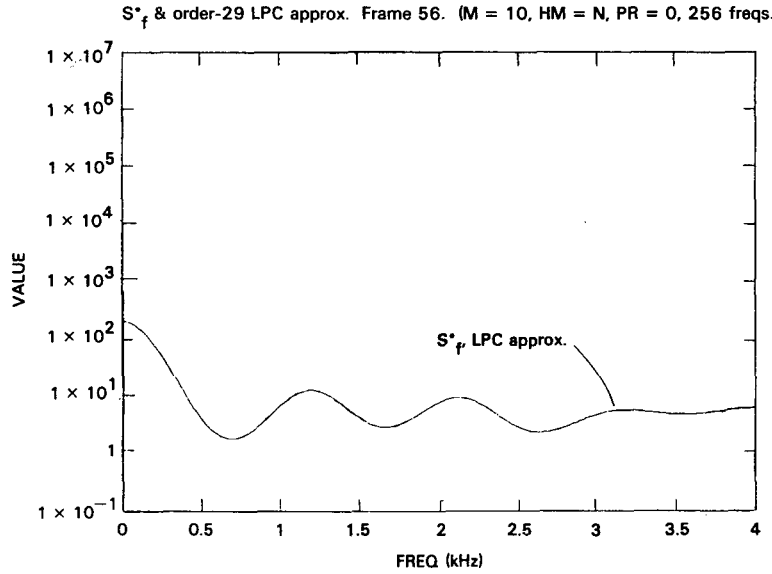

Fig. 3 — MENSA spectrum and the 29th-order continuous MESA approximation ($/f/$)

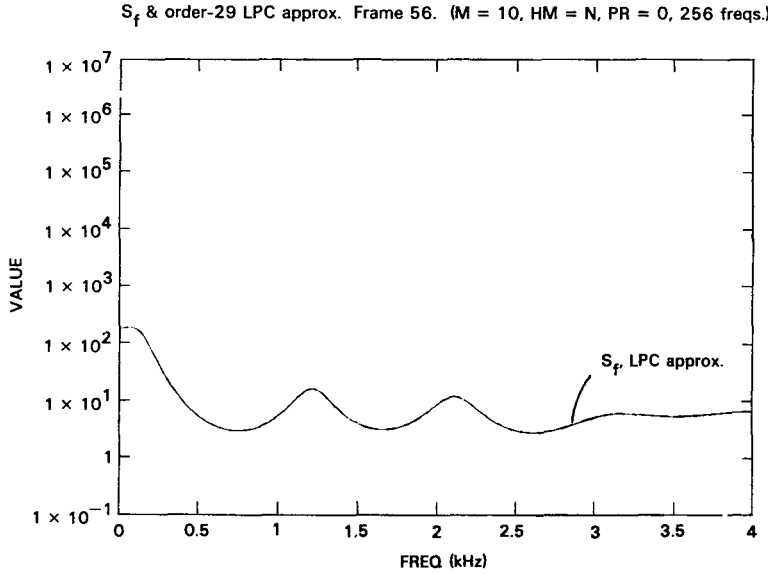

Fig. 4 — Discrete MESA spectrum and the 29th-order continuous MESA approximation ($/f/$)

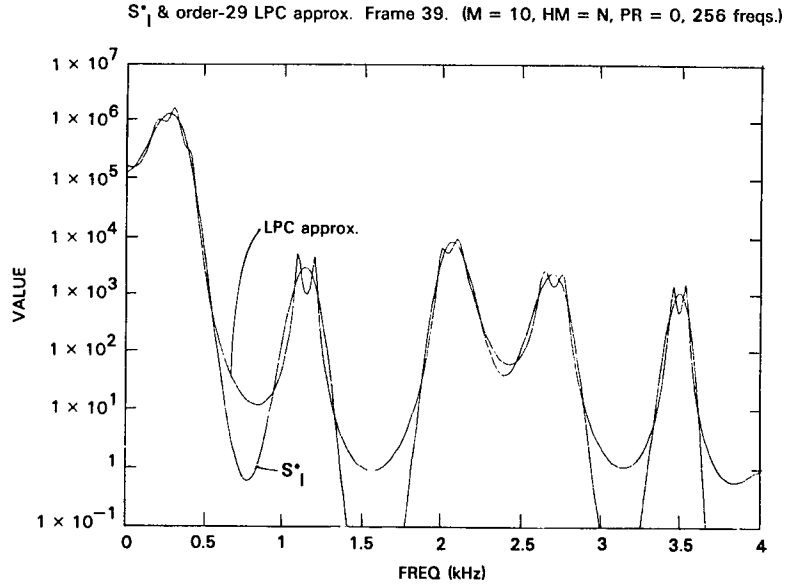

Fig. 5 — MENSA spectrum and the 29th-order continuous MESA approximation (*/I/*)

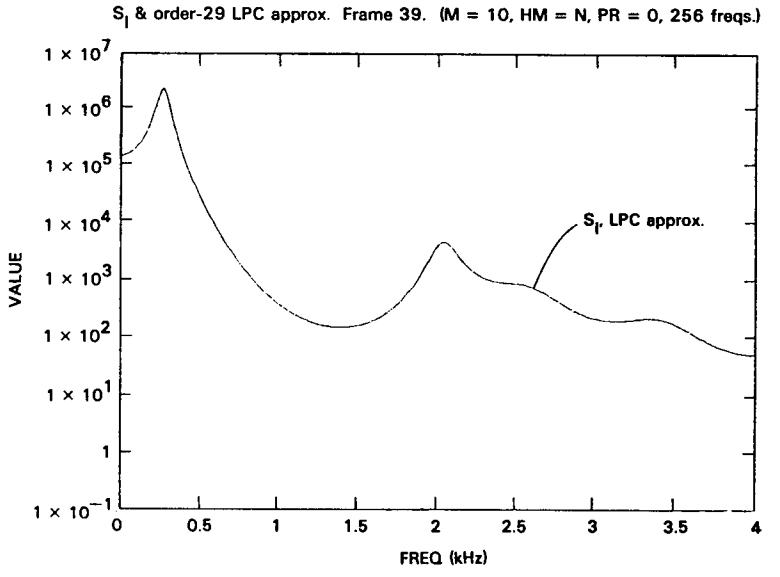

Fig. 6 — Discrete MESA spectrum and the 29th-order continuous MESA approximation (*/I/*)

are characteristic of the MENSA estimates for the entire test sentence. Indeed Fig. 7 shows the superimposed results of $\mathbf{S}^*$ for all 100 frames ($N = 256$). The frequent occurrence of five lobes is obvious. No such structure occurs for $\mathbf{S}$ (Fig. 8). The lobe structure appears to be related to the number of constraints: There are five lobes in Fig. 7, which is half the analysis order ($M = 10$). We repeated the computation of $\mathbf{A}^*$ using $M = 25$ and $M = 8$. The resulting plots were similar to Fig. 7 except that about 12 and four lobes were apparent respectively. Neither preemphasis nor windowing was entirely effective in eliminating the deep minima from the MENSA spectra. The superposed plots continued to show a lobed structure, though more complex and less regular than the consistent five-lobe pattern of Fig. 7. The results of using both Hamming windowing and 90% preemphasis are shown in Fig. 9. The lobes at 400 Hz, 1200 Hz, and 3600 Hz are still apparent, but the pattern is blurred between 2000 Hz and 3800 Hz.

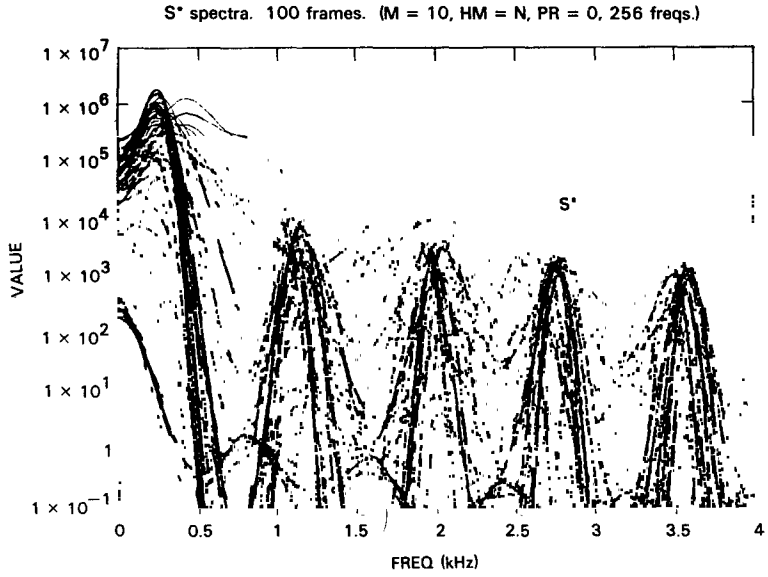

Fig. 7 — MENSA spectra—100 frames overlaid

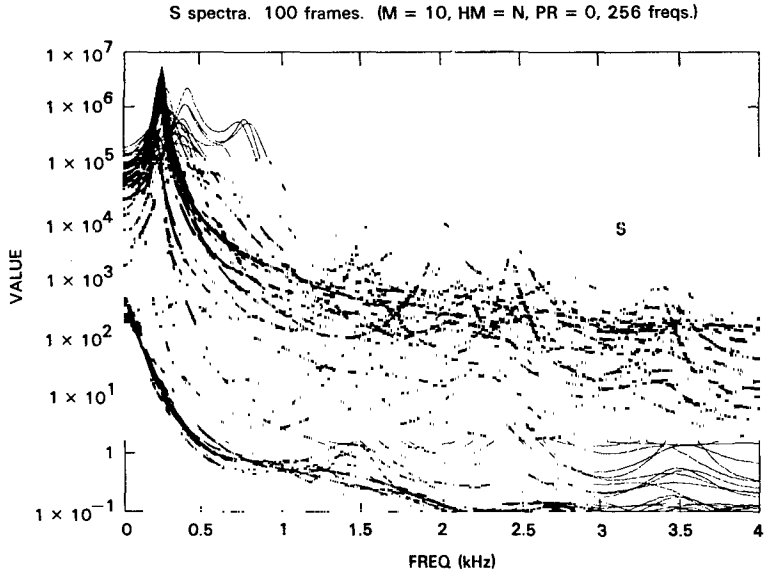

Fig. 8 — MESA spectra—100 frames overlaid

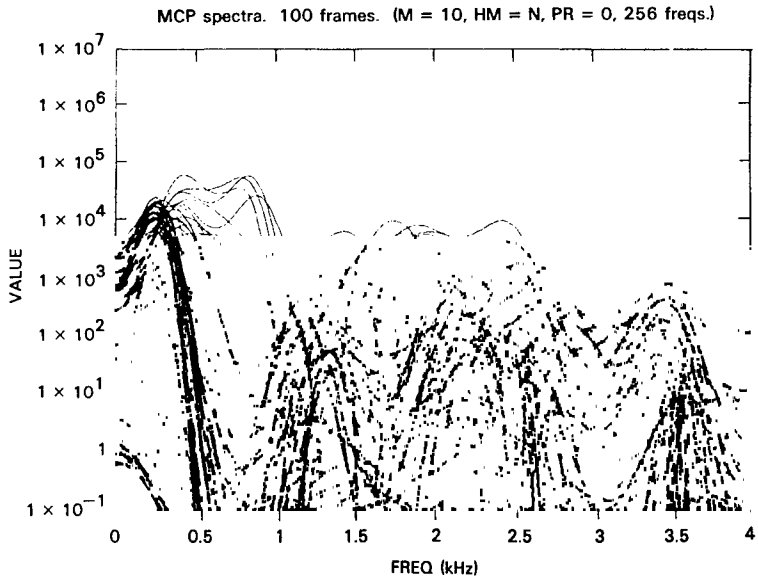

Fig. 9 — MENSA spectra from windowed, preemphasized speech— 100 frames overlaid

In Fig. 10 we compare the "actual" power spectrum $\mathbf{Q}_f$ with $\mathbf{S}_f^*$ and $\mathbf{S}_f$. Both estimates appear to be smoothed versions of $\mathbf{Q}_f$. Figure 11 shows the analogous comparison for /*I*/. Here there is more of a difference. Because of the deep minima of $\mathbf{S}^*$, it appears more reasonable to interpret $\mathbf{S}_I$ than $\mathbf{S}_I^*$ as a smoothed version of $\mathbf{Q}_I$.

Three distortion measures for the MESA and MENSA spectra $\mathbf{S}$ and $\mathbf{S}^*$ as estimates of $\mathbf{Q}$ were computed for each frame: $d_{IS}(\mathbf{Q},\mathbf{S})$ and $d_{IS}(\mathbf{Q},\mathbf{S}^*)$, $d_{GO}(\mathbf{Q},\mathbf{S})$ and $d_{GO}(\mathbf{Q},\mathbf{S}^*)$, and $d_{CE}(\mathbf{Q},\mathbf{S})$ and $d_{CE}(\mathbf{Q},\mathbf{S}^*)$. The computations were done for three values of M. The results, averaged over all 100 frames, are shown in Table 1. In one case the mean distortion for MENSA is slightly less than that for MESA, the difference being in the third decimal place. In every other case the mean distortion for MESA is less. This is true even for the "cross-entropy" distortions d_{CE}, which might have been expected to favor MENSA.

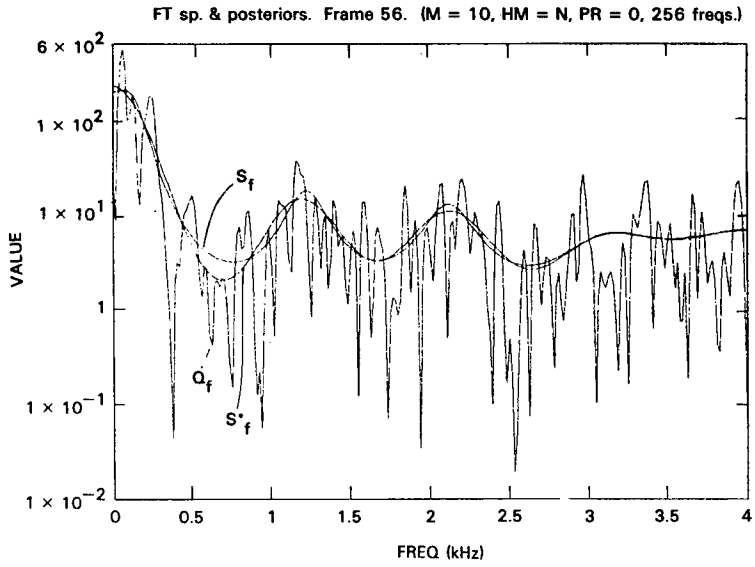

Fig. 10 — MESA and MENSA estimates with the Fourier transform of the measured autocorrelations (/*f*/)

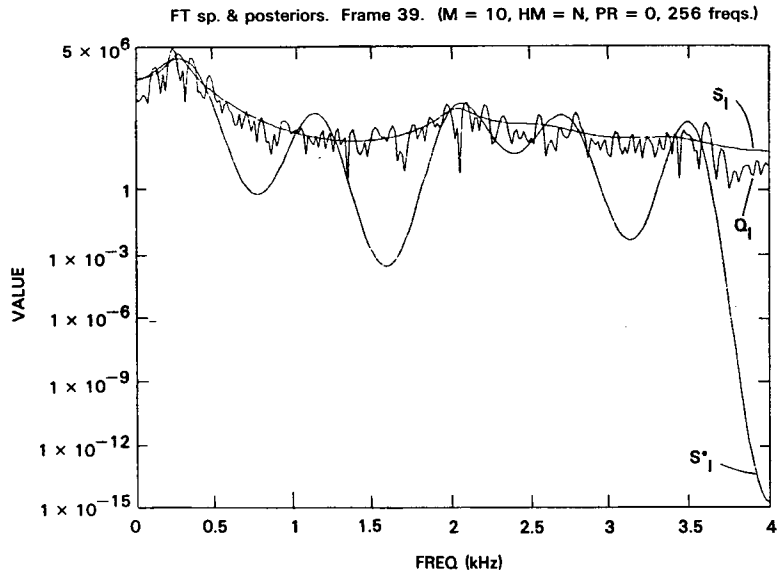

Fig. 11 — MESA and MENSA estimates with the Fourier transform of the measured autocorrelations ($/I/$)

Table 1 — Distortion Results

			Itakura-Saito Distortion d_{IS}		Gain-Optimized Itakura-Saito Distortion d_{GO}		Cross-Entropy Distortion d_{CE}	
M	Preemphasis	Window	MESA	MENSA	MESA	MENSA	MESA	MENSA
8	0	None	0.320	1.6×10^{20}	0.320	9.970	0.361	0.570
10	0	None	0.275	3.7×10^{19}	0.275	10.549	0.307	0.495
25	0	None	0.204	5.6×10^{18}	0.204	4.352	0.185	0.290
8	90%	Hamming	0.589	6.1×10^{18}	0.589	12.340	0.343	0.507
10	90%	Hamming	0.502	3.3×10^{18}	0.502	11.025	0.292	0.429
25	90%	Hamming	0.310	2.5×10^{17}	0.310	4.776	0.167	0.162
8	90%	None	0.434	6.5×10^{15}	0.434	5.520	0.426	0.596
10	90%	None	0.379	1.1×10^{19}	0.379	5.019	0.359	0.521
25	90%	None	0.265	1.1×10^{2}	0.265	1.032	0.194	0.262
8	0	Hamming	0.553	1.6×10^{20}	0.553	15.265	0.354	0.374
10	0	Hamming	0.446	8.7×10^{19}	0.446	16.347	0.243	0.321
25	0	Hamming	0.290	4.2×10^{19}	0.290	13.203	0.134	0.164

The d_{CE} results certainly do not favor MESA as overwhelmingly as those from the other two distortion measures—especially d_{IS}. The enormous Itakura-Saito distortions of the MENSA spectra are the result of the deep minima of the MENSA estimates. The expression for $d_{IS}(\mathbf{Q}, \mathbf{S}^*)$ contains the term Q_k/S_k^*, which becomes extremely large when the estimate S_k^* is nearly zero. The other two distortion measures contain such a term only logarithmically. Thus d_{IS} penalizes underestimates more severely than do d_{GO} and d_{CE}.

Two columns of the table are identical: it appears that $d_{IS}(\mathbf{Q}, \mathbf{S}) = d_{GO}(\mathbf{Q}, \mathbf{S})$. This is not a coincidence but is a property of d_{IS} and d_{GO}. The equality can be shown to hold provided that bS is a MESA spectrum and that bQ is a spectrum that satisfies the same autocorrelation constraints that determine bS. A proof can be based on the "correlation matching" property [39,29] of MESA spectra.

Comparison of Synthetic Speech

Although results such as in Figs. 7 and 11 suggest that $\mathbf{S}$ is better than $\mathbf{S}^*$, the separation is hardly compelling. This is a case where the proof must be in the hearing. Consequently we synthesized the entire test sentence using standard LPC methods and using the 29th-order LPC approximations to $\mathbf{S}_j^*$ and $\mathbf{S}_j$, $j = 1, \ldots, 100$, as discussed at the end of the preceding main section. The 29th-order LPC approximations to $\mathbf{S}_f^*$, $\mathbf{S}_f$, $\mathbf{S}_I^*$, and $\mathbf{S}_I$ are also plotted in Figs. 3 through 6. The two curves are indistinguishable in Figs. 3, 4, and 6; the only discrepancy is for S_I^* (Fig. 5). In that case the 29th-order approximation is unable to match the deep nulls and also exhibits some peak splitting.

The LPC speech and the speech based on $\mathbf{S}$ sounded identical, adding further confidence to the discrete frequency approximations. The two versions based on $\mathbf{S}$ and $\mathbf{S}^*$ sounded different, but—somewhat to our surprise—we and others judged them to be equally intelligible. There was, however, a distinct qualitative difference when preemphasis was not used. The speech based on $\mathbf{S}^*$ was qualitatively inferior—it had a distinct ringing quality, as though spoken from the other end of a long, wide pipe. When preemphasis was used, alone or with Hamming windowing, the ringing quality was greatly reduced or effectively eliminated. Hamming windowing alone reduced the ringing only slightly. We hypothesize that this ringing effect is a reflection of the characteristic lobe structure and deep minima of the spectral estimates $\mathbf{S}^*$, since the ringing is most prominent when the lobing is most prominent and regular. However, the ringing can be almost imperceptible while lobing is still plainly visible in spectral plots.

CONCLUSION

Primarily on the basis of the results of speech synthesis, but also on results like Fig. 7, Fig. 11, and Table 1, we believe that MESA ($\mathbf{S}$) yields better power spectrum estimates for speech processing than does MENSA ($\mathbf{S}^*$). This empirical conclusion also supports the theoretical discussion in the last paragraph of the background section.

REFERENCES

1. J.P. Burg, "Maximum entropy spectral analysis," presented at the 37th Annual Meeting, Soc. of Exploration Geophysicists, Oklahoma City, 1967.

2. J.P. Burg, "Maximum Entropy Spectral Analysis," Ph.D. dissertation, Stanford University, 1975 (University Microfilms 75-25, 449).

3. A. VanDenBos, "Alternative interpretation of maximum entropy spectral analysis, "IEEE Trans. Inf. Theory **IT-17**, 493-494 (July 1971).

4. J.D. Markel and A.H. Gray, Jr., *Linear Prediction of Speech*, Springer-Verlag, New York, 1976.

5. S.M. Kay and S.L. Marple, Jr., "Spectrum analysis—a modern perspective," Proc. IEEE **69**, 1380-1419 (Nov. 1981).

6. A. Papoulis, "Maximum entropy and spectral estimation: a review," IEEE Trans. Acoust., Speech, Signal Processing **ASSP-29**, 1176-1186 (Dec. 1981).

7. R.T. Lacoss, "Data adaptive spectral analysis methods," Geophysics **36**, 661-675 (1971).

8. T.J. Ulrych and T.N. Bishop, "Maximum entropy spectral analysis and autoregressive decomposition," Rev. Geophys. Space Phys. **43**, 183-200 (1975).

9. D.E. Smylie, G.K.C. Clarke, and T.J. Ulrych, "Analysis of irregularities in the earth's rotation," pp. 391-431 in *Methods in Computational Physics*, Vol. **13**, Academic Press, New York, 1973.

10. R.M. Gray, A.H. Gray, Jr., G. Rebolledo, and J.E. Shore, "Rate-distortion speech coding with a minimum discrimination information distortion measure," IEEE Trans. Inform. Theory **IT-27**, 708-721 (Nov. 1981).

11. S.J. Wernecke and L. D'Addario, "Maximum entropy image reconstruction," IEEE Trans. Computers **C-26**(4), 351-364 (Apr. 1977).

12. J.G. Ables, "Maximum entropy spectral analysis," Astron. Astrophys. Suppl. **15**, 383-393 (1974).

13. S.J. Wernecke, "Two-dimensional maximum entropy reconstruction of radio brightness," Radio Science **12**(5), 831-844 (1977).

14. B.R. Frieden, "Restoring with maximum likelihood and maximum entropy," J. Opt. Soc. Am. **62**(4), 511-518 (Apr. 1972).

15. R. Gordon and G.T. Herman, "Reconstruction of pictures from their projections," Quarterly Bull. Center for Theor. Biol. **4**, 71-151 (1971).

16. J. Skilling, "Maximum entropy and image processing—algorithms and applications," in *Proceedings, First Maximum Entropy Workshop*, 1981.

17. M.D. Ortigueira, R. Garcia-Gomez, and J.M. Tribolet, "An iterative algorithm for maximum flatness spectral analysis," pp. 810-818 in *Proceedings, International Conference on DSP*, 1981.

18. C. Nadeu, E. Sanvicente, and M. Bertran, "A new algorithm for spectral estimation," pp. 463-470 in *Proceedings, International Conference on DSP*, 1981.

19. R. Kikuchi and B.H. Soffer, "Maximum entropy image restoration. I. The entropy expression," J. Opt. Soc. Am. **67**(12), 1656-1665 (1977).

20. E.T. Jaynes, "Information theory and statistical mechanics I," Phys. Rev. **106**, 620-630 (1957).

21. E.T. Jaynes, "Information theory and statistical mechanics II," Phys. Rev. **108**, 171-190 (1957).

22. W.M. Elsasser, "On quantum measurements and the role of the uncertainty relations in statistical mechanics," Phys. Rev. **52**, 987-999 (Nov. 1937).

23. C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, Univ. Illinois Press, Chicago, 1949.

24. M.S. Bartlett, *An Introduction to Stochastic Processes*, Cambridge Univ. Press, Cambridge, England, 1966.

25. J.E. Shore, "Minimum cross-entropy spectral analysis," IEEE Trans. Acoust., Speech, Signal Processing **ASSP-29**, 230-237 (Apr. 1981).

26. R.G. Gallagher, *Information Theory and Reliable Communication*, Wiley, New York, 1968.

27. S. Kullback, *Information Theory and Statistics*, Dover, New York, 1969, and Wiley, New York, 1959.

28. J.E. Shore and R.W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," IEEE Trans. Inform. Theory **IT-26**, 26-37 (Jan. 1980).

29. J.E. Shore and R.W. Johnson, "Properties of cross-entropy minimization," IEEE Trans. Inform. Theory **IT-27**, 472-482 (July 1981).

30. E.T. Jaynes, "Prior probabilities," IEEE Trans. Systems Science and Cybernetics **SSC-4**, 227-241 (1968).

31. I. Csiszár, "I-divergence geometry of probability distributions and minimization problems," Ann. Math. Stat. **3**, 146-158 (1975).

32. Y. Yaglom, *An Introduction to the Theory of Stationary Random Functions*, Prentice-Hall, Englewood Cliffs, N.J., 1962.

33. J.E. Shore, "Minimum Cross-Entropy Spectral Analysis," NRL Memorandum Report 3921, Jan. 1979.

34. L. Vergara-Domínguez and A.R. Figueiras-Vidal, "A minimum cross-flatness spectral estimator and some related problems," preprint, private communication.

35. C.L. Byrne and R.M. Fitzgerald, "Reconstruction from partial information with applications to tomography," SIAM J. Appl. Math. **42**(4), 933-940 (Aug. 1982).

36. R.M. Fitzgerald, private communication.

37. F. Itakura and S. Saito, "Analysis synthesis telephony based on the maximum likelihood method," in *Reports of the 6th International Congress on Acoustics*, 1968.

38. F. Itakura and S. Saito, "A statistical method for estimation of speech spectral density and formant frequencies," Electron. Commun. Japan **53-A**, 36-43 (1970).

39. R.M. Gray, A. Buzo, A.H. Gray, Jr., and Y. Matsuyama, "Distortion measures for speech processing," IEEE Trans. Acoust., Speech, Signal Processing **ASSP-28**, 367-376 (Aug. 1980).

40. R.W. Johnson, "Determining probability distributions by maximum entropy and minimum cross-entropy," in *APL79 Conference Proceedings*, ACM 0-89791-005 (May 1979).

41. R.W. Johnson, "Algorithms for single-signal and multisignal minimum-cross-entropy spectral analysis," NRL Report 8667, 1983; also to be submitted to IEEE Trans. Acoustics, Speech, Signal Proc.

42. R.W. Schafer and J.D. Markel, editors, *Speech Analysis*, IEEE Press, New York (1979).